Linear Algebra and its Applications 434 (2011) 1425–1443

Contents lists available at ScienceDirect

Linear Algebra and its Applications

journal homepage: www.elsevier.com/locate/laa

On entropy-preserving stochastic averages

Alan B. Poritz [a], Jonathan A. Poritz [b,*]

[a] Center for Communications Research – Princeton, 805 Bunn Drive, Princeton, NJ 08540, USA
[b] Department of Mathematics and Physics, Colorado State University, Pueblo, 2200 Bonforte Blvd., Pueblo, CO 81001, USA

ARTICLE INFO

Article history:
Received 5 June 2010
Accepted 18 October 2010
Available online 18 November 2010

Submitted by R. Guralnick

AMS classification:
94A15
05C50

Keywords:
Entropy
Stochastic average
Doubly stochastic matrix
Birkhoff polytope
Probability simplex

ABSTRACT

When an $n \times n$ doubly stochastic matrix A acts on $\mathbb{R}^n$ on the left as a linear transformation and P is an n-long probability vector, we refer to the new probability vector AP as the *stochastic average* of the pair (A, P). Let $\mathbf{\Gamma}_n$ denote the set of pairs (A, P) whose stochastic average preserves the entropy of P: $H(AP) = H(P)$. $\mathbf{\Gamma}_n$ is a subset of $\mathbf{B}_n \times \mathbf{\Sigma}_n$ where $\mathbf{B}_n$ is the Birkhoff polytope and $\mathbf{\Sigma}_n$ is the probability simplex.
We characterize $\mathbf{\Gamma}_n$ and determine its geometry, topology,and combinatorial structure. For example, we find that $(A, P) \in \mathbf{\Gamma}_n$ if and only if $A^tAP = P$. We show that for any n, $\mathbf{\Gamma}_n$ is a connected set, and is in fact piecewise-linearly contractible in $\mathbf{B}_n \times \mathbf{\Sigma}_n$. We write $\mathbf{\Gamma}_n$ as a finite union of product subspaces, in two distinct ways. We derive the geometry of the fibers $(A, \cdot)$ and $(\cdot, P)$ of $\mathbf{\Gamma}_n$. $\mathbf{\Gamma}_3$ is worked out in detail. Our analysis exploits the convexity of $x \log x$ and the structure of an efficiently computable bipartite graph that we associate to each ds-matrix. This graph also lets us represent such a matrix in a permutation-equivalent, block diagonal form where each block is doubly stochastic and fully indecomposable.

1. Introduction

In his 1948 paper *A Mathematical Theory of Communications* [1,2], Shannon introduced entropy as a measure of information in a probability distribution. If the column vector $P = (p_1, \ldots, p_n)^t \in \mathbb{R}^n$ is a probability distribution, he defined its entropy to be the quantity

* Corresponding author. Tel.: +1 719 337 1210; fax: +1 719 549 2962.
E-mail address: alan.poritz@gmail.com (A.B. Poritz), jonathan.poritz@gmail.com (J.A. Poritz).
URL: http://www.poritz.net/jonathan (J.A. Portitz)

0024-3795/$ - see front matter
doi:10.1016/j.laa.2010.10.014

$$H = H(P) = -\sum_{i=1}^{n} p_i \log_2(p_i),$$

where $p \log_2(p)$ is set to 0 whenever $p = 0$. He remarked, [1, p. 395]:

> "Any change toward equalization of the probabilities $p_1, p_2, \ldots, p_n$ increases H. Thus if $p_1 < p_2$ and we increase p_1, decreasing p_2 by an equal amount so that p_1 and p_2 are more nearly equal, H increases. More generally, if we perform any "averaging" operation on the p_i of the form
>
> $$p'_i = \sum_j a_{ij} p_j$$
>
> where $\sum_i a_{ij} = \sum_j a_{ij} = 1$ and all $a_{ij} \geqslant 0$, then H increases (except in the special case where this transformation amounts to no more than a permutation of the p_j with H of course remaining the same)."

Let $A = (a_{ij})$ denote the $n \times n$ doubly stochastic matrix (*ds-matrix* for short) given above. We refer to the probability distribution $P' = (p'_1, \ldots, p'_n)^t = AP$ as the *stochastic average of the pair* (A, P). In these terms, what Shannon has said is that a necessary and sufficient condition for a stochastic average to preserve the entropy of P (*i.e.*, $H(AP) = H(P)$) is that there exists an $n \times n$ permutation matrix π such that $AP = \pi P$.

This viewpoint provokes several questions. What is the structure of the set of all ds-matrices A whose stochastic average with a fixed P preserves its entropy? What is the structure of the set of all P whose stochastic average with a fixed ds-matrix A preserves the entropy of P? And, finally, what is the structure of the set of all pairs (A, P) whose stochastic averages preserve entropy? In what follows, we present a number of criteria for membership in these sets and then explore their geometry, topology and combinatorics.

We write $\mathbf{S}_n \subset \mathbb{R}^{n^2}$ for the set of $n \times n$ permutation matrices, and identify the permutation group on $\{1, \ldots, n\}$ with $\mathbf{S}_n$: a permutation takes i into j if and only if the 1 in the ith column of the corresponding matrix occurs in its jth row. Hence a permutation of $\{1, \ldots, n\}$ carries i into j if and only if left multiplication by the corresponding permutation matrix carries $\mathbf{e}^n_i$ into $\mathbf{e}^n_j$, where $\{\mathbf{e}^n_1, \ldots, \mathbf{e}^n_n\}$ is the standard basis of $\mathbb{R}^n$.

The convex hull of $\mathbf{S}_n$ in $\mathbb{R}^{n^2}$ is called the *Birkhoff polytope* and denoted $\mathbf{B}_n$. The Birkhoff–von Neumann Theorem (*e.g.*, [3]) states that $\mathbf{B}_n$ is an $(n-1)^2$-dimensional polytope with $n!$ vertices and consists of the set of all $n \times n$ ds-matrices.

Example 1.1. The set of 2×2 ds-matrices is $\mathbf{B}_2 = \left\{ \begin{pmatrix} t & 1-t \\ 1-t & t \end{pmatrix} \;\middle|\; 0 \leqslant t \leqslant 1 \right\}$.

This set forms a line segment in $\mathbb{R}^4$.

Let $\mathbf{\Sigma}_n \subset \mathbb{R}^n$ denote the set of n-long probability distributions. $\mathbf{\Sigma}_n$ is an $(n-1)$-dimensional simplex, the convex hull of $\{\mathbf{e}^n_1, \ldots, \mathbf{e}^n_n\}$. $\mathbf{\Sigma}_n$ contains the *uniform distribution* $U = (1/n, \ldots, 1/n)^t$; U uniquely maximizes the entropy function H on $\mathbf{\Sigma}_n$, [1].

Definition 1.2. The set of *entropy-preserving stochastic averages* is

$$\mathbf{\Gamma}_n \triangleq \{(A, P) \in \mathbf{B}_n \times \mathbf{\Sigma}_n \mid H(AP) = H(P)\}.$$

$\mathbf{\Gamma}_n$ contains both $\mathbf{S}_n \times \mathbf{\Sigma}_n$ and $\mathbf{B}_n \times \{U\}$. Notice also that if $A \in \mathbf{B}_n$ is block diagonal and $P \in \mathbf{\Sigma}_n$ is constant on the indices in each block, then $(A, P) \in \mathbf{\Gamma}_n$.

Further, there is a natural left action of $\mathbf{S}_n \times \mathbf{S}_n$ on $\mathbf{B}_n \times \mathbf{\Sigma}_n$ given by

$$(\mathbf{S}_n \times \mathbf{S}_n) \times (\mathbf{B}_n \times \mathbf{\Sigma}_n) \to \mathbf{B}_n \times \mathbf{\Sigma}_n : ((\pi_1, \pi_2), (A, P)) \mapsto (\pi_1 A \pi_2^{-1}, \pi_2 P) \tag{1.1}$$

and we have:

Proposition 1.3. *$\mathbf{\Gamma}_n$ is invariant under the action of $\mathbf{S}_n \times \mathbf{S}_n$ on $\mathbf{B}_n \times \mathbf{\Sigma}_n$.*

Proof. Start with $(A, P) \in \mathbf{\Gamma}_n$, so that $H(AP) = H(P)$, and any $\pi_1, \pi_2 \in \mathbf{S}_n$. Then

$$H((\pi_1 A \pi_2^{-1})(\pi_2 P)) = H(\pi_1 AP) = H(AP) = H(P) = H(\pi_2 P)$$

so $(\pi_1, \pi_2) \cdot (A, P) = (\pi_1 A \pi_2^{-1}, \pi_2 P) \in \mathbf{\Gamma}_n$. $\square$

To reach a more complete understanding of $\mathbf{\Gamma}_n$ we introduce in Section 2 a pair of partitions of $\{1, \ldots, n\}$ for each $(A, P) \in \mathbf{\Gamma}_n$. One partition is induced on the domain of P by its level sets. For the other, we show that the positive entries of a ds-matrix $A \in \mathbf{B}_n$ determine a bipartite graph whose connected components induce a partition on the set of column indices of A. In Section 4 we prove that $(A, P) \in \mathbf{\Gamma}_n$ if and only if the partition induced on the column indices of A refines the partition induced on the domain of P. We then give a proof of Shannon's interpretation of $\mathbf{\Gamma}_n$ as consisting of those pairs (A, P) satisfying $AP = \pi P$ for some permutation π. And we establish a third characterization of $\mathbf{\Gamma}_n$: $(A, P) \in \mathbf{\Gamma}_n$ if and only if $A^t AP = P$.

The characterization of $\mathbf{\Gamma}_n$ in terms of refinement of partitions gives us access to both its global properties and its fine structure. In Section 5, we show that a fiber of $\mathbf{\Gamma}_n$ of the form $(A, \cdot)$ is both a sub-simplex of $\mathbf{\Sigma}_n$ and a sub-complex of its barycentric subdivision [4]. In Section 6, we observe that a fiber of the form $(\cdot, P)$ is a union of disjoint and isomorphic faces of the polytope $\mathbf{B}_n$. That the set of fibers of either form can be placed in bijective correspondence with the set of the partitions of $\{1, \ldots, n\}$ is shown in Section 8. We show in Section 7 that $\mathbf{\Gamma}_n$ is a contractible and hence connected set. Finally, in Section 9, we establish a pair of partitions of $\mathbf{\Gamma}_n$ into finitely many product subsets and then explicitly compute these partitions for $\mathbf{\Gamma}_2$ and $\mathbf{\Gamma}_3$.

One could develop a structure, entirely parallel to what we present here, to describe the set of pairs (A, P) where entropy is preserved when a ds-matrix A acts on the right on a probability distribution $P : P^t \mapsto P^t A$.

2. Partitions and decompositions

2.1. Set partitions and integer partitions

Let $\mathbb{N}$ denote the set of natural numbers. We make frequent use of both partitions of the set $\{1, \ldots, n\} \subset \mathbb{N}$ and partitions of the integer $n \in \mathbb{N}$.

A *partition of* $\{1, \ldots, n\}$ is a set of non-empty, pairwise disjoint subsets whose union is all of $\{1, \ldots, n\}$; the collection of all partitions of $\{1, \ldots, n\}$ will be denoted $\mathbf{\Phi}_n$. The subset of $\mathbf{\Phi}_n$ consisting of those set partitions with exactly m subsets will be denoted $\mathbf{\Phi}_{n,m}$. While set partitions are intrinsically unordered, if $\mathcal{J} \in \mathbf{\Phi}_{n,m}$ is indexed as $\mathcal{J} = \{J_1, \ldots, J_m\}$, where the sequence of least elements of the J_i's increases as i goes from 1 up to m, we say that $\mathcal{J}$ *is indexed in natural order*.

A *partition of* $n \in \mathbb{N}$ is a set of positive integers whose sum is n; the collection of all partitions of n will be denoted ϕ_n. The subset of ϕ_n consisting of all integer partitions with exactly m summands will be denoted $\phi_{n,m}$.

The two uses of the word "partition" are closely related.

Definition 2.1. Let $\mathcal{J} \in \mathbf{\Phi}_{n,m}$ be a partition of the set $\{1, \ldots, n\}$ and let $\mathfrak{n} \in \phi_{n,m}$ be the partition of the integer n whose summands are the cardinalities of the subsets in $\mathcal{J}$. We refer to $\mathfrak{n}$ as the *integer partition induced by* $\mathcal{J}$.

If the set partition $\mathcal{J}$ is ordered, then the integer partition $\mathfrak{n}$ inherits this ordering.

Definition 2.2. Let $\mathfrak{n} = \{n_1, \ldots, n_m\} \in \phi_{n,m}$ be an ordered partition of the integer n and let $\mathcal{N} = \{N_1, \ldots, N_m\} \in \mathbf{\Phi}_{n,m}$ be the ordered set partition such that $N_1 = \{1, \ldots, n_1\}$ and, if $m > 1$, each successive N_i consists of the next n_i elements of $\{1, \ldots, n\}$, for $i \in \{2, \ldots, m\}$. We refer to $\mathcal{N}$ as the *set partition induced by* $\mathfrak{n}$.

The induced partition $\mathcal{N}$ is in natural order.

Suppose that the set partition $\mathcal{J} \in \mathbf{\Phi}_{n,m}$ is ordered; say $\mathcal{J} = \{J_1, \ldots, J_m\}$. Then both the induced integer partition $\mathfrak{n} = \{n_1, \ldots, n_m\}$ and the subsequent induced set partition $\mathcal{N} = \{N_1, \ldots, N_m\}$ are ordered. The cardinality of J_i and the cardinality of N_i both equal n_i. Thus there is a permutation $\pi_{\mathcal{J}} \in \mathbf{S}_n$ uniquely defined by the requirement that it carries the kth smallest integer in J_i to the kth smallest integer in N_i for $k = 1, \ldots, n_i$ and $i = 1, \ldots, m$.

Definition 2.3. We refer to $\pi_{\mathcal{J}}$ as the *unscrambling permutation* induced by the ordered partition $\mathcal{J}$.

Note that the action of the permutation group $\mathbf{S}_n$ on $\{1, \ldots, n\}$ extends in a natural way to an action on the set of partitions $\mathbf{\Phi}_n$ preserving $\mathbf{\Phi}_{n,m}$: if $\pi \in \mathbf{S}_n$ and $\mathcal{J} \in \mathbf{\Phi}_{n,m}$ then $\pi(\mathcal{J}) \in \mathbf{\Phi}_{n,m}$ is the partition such that $J' \in \pi(\mathcal{J})$ if and only if there is some $J \in \mathcal{J}$ such that $J' = \pi(J) = \{\pi(j) | j \in J\}$.

2.2. Going from partitions to ds-matrices and probability distributions

Partitions generate distinguished subsets of the Birkhoff polytope and the probability simplex.

Definition 2.4. Fix a set partition $\mathcal{J} \in \mathbf{\Phi}_n$.

1. The *$\mathcal{J}$-adapted permutation matrices* are the members of the set
   $$\mathbf{S}_n(\mathcal{J}) \triangleq \left\{\pi \in \mathbf{S}_n | \text{ if } \pi_{ij} \neq 0 \text{ then there is a } J \in \mathcal{J} \text{ with both } i, j \in J\right\};$$
   i.e., elements of $\mathbf{S}_n(\mathcal{J})$ permute the indices of each component of $\mathcal{J}$ separately.
2. The *$\mathcal{J}$-adapted ds-matrices* $\mathbf{B}_n(\mathcal{J})$ are the convex hull of $\mathbf{S}_n(\mathcal{J})$ in $\mathbb{R}^{n^2}$.
3. Similarly, the *$\mathcal{J}$-adapted probability distributions* are the members of the set
   $$\mathbf{\Sigma}_n(\mathcal{J}) \triangleq \left\{P \in \mathbf{\Sigma}_n | \text{ if } p_i \neq p_j \text{ then there is a } J \in \mathcal{J} \text{ with } i \in J, j \notin J\right\};$$
   i.e., $\mathbf{\Sigma}_n(\mathcal{J})$ consists of probability distributions which are constant on the indices in each $J \in \mathcal{J}$.
4. Now fix an ordered partition $\mathfrak{n} = \{n_1, \ldots, n_m\} \in \phi_{n,m}$ and set
   $$\mathbf{S}_n(\mathfrak{n}) \triangleq \mathbf{S}_n(\mathcal{N}), \quad \mathbf{B}_n(\mathfrak{n}) \triangleq \mathbf{B}_n(\mathcal{N}), \quad \text{and} \quad \mathbf{\Sigma}_n(\mathfrak{n}) \triangleq \mathbf{\Sigma}_n(\mathcal{N}),$$
   where $\mathcal{N} \in \mathbf{\Phi}_{n,m}$ is the set partition induced by $\mathfrak{n}$.

$\mathbf{S}_n(\mathfrak{n})$ and $\mathbf{B}_n(\mathfrak{n})$ are each a set of block diagonal matrices in $M_{n\times n}(\mathbb{R})$ where block i has size $n_i \times n_i$ for $i = 1, \ldots, m$. It further follows from their definitions that

$$\mathbf{S}_n(\mathfrak{n}) \cong \mathbf{S}_{n_1} \times \cdots \times \mathbf{S}_{n_m} \quad \text{and} \quad \mathbf{B}_n(\mathfrak{n}) \cong \mathbf{B}_{n_1} \times \cdots \times \mathbf{B}_{n_m}.$$

In particular, $\mathbf{S}_n(\mathfrak{n})$ is a group.

2.3. Going from probability distributions and ds-matrices to partitions

To the level sets of a probability distribution in $\mathbf{\Sigma}_n$ there is naturally associated a set partition in $\mathbf{\Phi}_n$:

Definition 2.5. Given $P = (p_1, \ldots, p_n)^t \in \mathbf{\Sigma}_n$, let $\mathcal{I}_P \in \mathbf{\Phi}_n$ denote the partition of the domain of P which places i and j in the same subset if and only if $p_i = p_j$. We refer to $\mathcal{I}_P$ as the *coincidence partition of P*.

Note $\mathcal{I}_P \in \mathbf{\Phi}_{n,l}$, where l is the number of level sets of P. Notice also that if $\pi \in \mathbf{S}_n$, then $\mathcal{I}_{\pi P} = \pi(\mathcal{I}_P)$.

Next, we use a two-stage process to associate a set partition to a ds-matrix:

Definition 2.6. Given $A = (a_{ij}) \in \mathbf{B}_n$, define its *weight graph* to be the bipartite graph W_A whose vertices consist of two disjoint copies of $\{1, \ldots, n\}$ – we call these the *row indices* R and the *column indices* C, so the vertex set of W_A is $R \amalg C$ – and whose edges connect an $r \in R$ and a $c \in C$ if and only if the *weight* $a_{rc} > 0$.

The topology of the weight graph gives the desired partition:

Definition 2.7. Given a ds-matrix $A \in \mathbf{B}_n$, if its weight graph W_A has m connected components then we call m its *component count*. We denote these components $\{W_1, \ldots, W_m\}$, and define the *row-index partition* $\mathcal{R}_A = \{R_1, \ldots, R_m\} \in \mathbf{\Phi}_{n,m}$ by setting R_i to be the set of all row indices that appear in W_i, for each $i \in \{1, \ldots, m\}$, and the *column-index partition* $\mathcal{C}_A = \{C_1, \ldots, C_m\} \in \mathbf{\Phi}_{n,m}$ by setting C_i to be the set of all column indices that appear in W_i, for each $i \in \{1, \ldots, m\}$.

We index the partition $\mathcal{C}_A$ in natural order and then index the partitions $\mathcal{R}_A$ and W_A such that R_i and C_i come from the same component W_i, for each $i \in \{1, \ldots, m\}$.

The row- and column-index partitions are closely related:

Proposition 2.8. *Let A be a ds-matrix with weight graph W_A, component count m, row-index partition $\mathcal{R}_A = \{R_1, \ldots, R_m\}$, and column-index partition $\mathcal{C}_A = \{C_1, \ldots, C_m\}$. Then for each $i \in \{1, \ldots, m\}$,*

$$\#(R_i) = \#(C_i).$$

Proof. In any component of W_A, the sum of the weights on its edges equals the number of row indices in that component, since the sum of all weights going into any individual row vertex is 1. By the same reasoning, this sum must also equal the number of column indices in that component. Therefore the corresponding sets in $\mathcal{R}_A$ and $\mathcal{C}_A$ have the same cardinality. □

Definition 2.9. For a ds-matrix $A \in \mathbf{B}_n$ with weight graph W_A, component count m, row-index partition $\mathcal{R}_A = \{R_1, \ldots, R_m\} \in \mathbf{\Phi}_{n,m}$ and column-index partition $\mathcal{C}_A = \{C_1, \ldots, C_m\} \in \mathbf{\Phi}_{n,m}$, we call the partition $\mathfrak{n}_A = \{n_1, \ldots, n_m\} \in \phi_{n,m}$ of n defined either by $n_i = \#(R_i)$ or $n_i = \#(C_i)$, for $i \in \{1, \ldots, m\}$, the *component size partition of A*. We give $\mathfrak{n}_A$ the ordering induced by the natural order of the column-index partition $\mathcal{C}$.

When the matrix A in question is clear from context, we suppress the subscript A in all of the objects defined above: W instead of W_A, $\mathcal{C}$ instead of $\mathcal{C}_A$, *etc.* Likewise, if the probability distribution P is clear, we write $\mathcal{I}$ instead of $\mathcal{I}_P$.

Given a ds-matrix $A \in \mathbf{B}_n$, there is an efficient algorithm for constructing its weight graph and associated objects in m rounds. One connected component of the graph is produced in each round. At the start of the ith round, the component's sets of row indices R_i and edges E_i are both empty while the component's set of column indices C_i has as single entry: the smallest column index not yet used in earlier rounds. Each round consists of a sequence of steps; each step consists of two moves. Move 1: for each column index c just adjoined to C_i, adjoin to R_i the row indices for which A has positive entries in column c, and adjoin to E_i the corresponding edges. Move 2: for each row index r just adjoined to R_i, adjoin to C_i the column indices for which A has positive entries in row r, and adjoin to E_i the corresponding edges. After each step, test whether the cardinality of R_i equals the cardinality of C_i. If they are not equal, step again; if they are equal, the component W_i is complete and a new round begins. The process terminates in at most n steps. Pseudocode for this is shown as Algorithm 2.1.

Algorithm 2.1. Graph- and partition-creation algorithm

```
1: i ← 0
2: C^u ← {1, ..., n}
```

repeat
  $i \leftarrow i + 1$
  $R_i \leftarrow \emptyset$
  $C_i \leftarrow \min C^u$
  repeat
    $R_i \leftarrow R_i \cup \{r \in R | a_{rc} > 0 \text{ for some } c \in C_i\}$
    $C_i \leftarrow C_i \cup \{c \in C | a_{rc} > 0 \text{ for some } r \in R_i\}$
  until $\#(R_i) = \#(C_i)$
  define W_i to be the graph with vertices $R_i \amalg C_i$
    and edges $\{(r, c) \in R_i \times C_i | a_{rc} > 0\}$
  $n_i \leftarrow \#(R_i)$ (or $\#(C_i)$)
  $C^u \leftarrow C^u \setminus C_i$
until C^u is empty

output
- COMPONENT COUNT: $m \leftarrow i$
- WEIGHT GRAPH: $W \leftarrow \bigcup_{i=1}^{m} W_i$
- ROW-INDEX PARTITION: $\mathcal{R} = \{R_1, \ldots, R_m\}$
- COLUMN-INDEX PARTITION: $\mathcal{C} = \{C_1, \ldots, C_m\}$
- COMPONENT SIZE PARTITION: $\mathfrak{n} = \{n_1, \ldots, n_m\}$

2.4. A block diagonal equivalent form for doubly stochastic matrices

Recall that two $n \times n$ matrices A and B are called *permutation equivalent* if there exist $\pi_1, \pi_2 \in \mathbf{S}_n$ such that $B = \pi_1 A \pi_2^{-1}$. In other words, B is in the orbit of A under the left action $(\mathbf{S}_n \times \mathbf{S}_n) \times M_{n\times n}(\mathbb{R}) \to M_{n\times n}(\mathbb{R}) : ((\pi_1, \pi_2), A) \mapsto \pi_1 A \pi_2^{-1}$. (See, *e.g.*, [3] for further explication of this and other common terms used in this section).

The notion of permutation equivalence plays well with weight graphs and the partitions we have defined in Section 2.3:

Proposition 2.10. *If the matrices $A, B \in \mathbf{B}_n$ are permutation equivalent then their weight graphs are isomorphic, $W_B \cong W_A$. In fact, if $B = \pi_1 A \pi_2^{-1}$ for $\pi_1, \pi_2 \in \mathbf{S}_n$, then the row- and column-index partitions are related by $\mathcal{R}_B = \pi_1(\mathcal{R}_A)$ and $\mathcal{C}_B = \pi_2(\mathcal{C}_A)$, while the component size partitions are equal.*

An $n \times n$ matrix A is said to be *partly decomposable* if it contains an $m \times (n - m)$ zero submatrix for some m satisfying $0 < m < n$; that is, if there are permutations $\pi_1, \pi_2 \in \mathbf{S}_n$ such that

$$\pi_1 A \pi_2^{-1} = \begin{pmatrix} B & C \\ 0_{m\times(n-m)} & D \end{pmatrix}. \tag{2.1}$$

A is called *fully indecomposable* if it is not partly decomposable.

Here then is our equivalent form for ds-matrices:

Theorem 2.11. *Let $A \in \mathbf{B}_n$ have component size partition $\mathfrak{n} = \{n_1, \ldots, n_m\}$. Then A is permutation equivalent to a matrix in $\mathbf{B}_n(\mathfrak{n})$, each of whose m blocks is fully indecomposable.*

Proof. Say the row- and column-index partitions of A are $\mathcal{R} = \{R_1, \ldots, R_m\}$ and $\mathcal{C} = \{C_1, \ldots, C_m\}$, with $\mathcal{C}$ indexed in its natural order and $\mathcal{R}$ indexed so that R_i and C_i come from the same component of the weight graph W_A, for $i \in \{1, \ldots, m\}$. Then Proposition 2.10 tells us immediately that the permutation equivalence using the unscrambling permutations $\pi_2 = \pi_{\mathcal{C}}$ and $\pi_1 = \pi_{\mathcal{R}}$ has exactly the desired form in $\mathbf{B}_n(\mathfrak{n})$.

To show the blocks are fully indecomposable, we restrict our attention to a single block in $\mathbf{B}_n(\mathfrak{n})$. Here it suffices to show that if a matrix A is partly decomposable, then its component count must be greater than 1. Let notation be as above at Eq. (2.1). Since $\pi_1 A \pi_2^{-1}$ is column-stochastic, the sum of all entries in the first $n - m$ columns must be $n - m$, which must then be the sum of all the elements of B as the block of zeros contributes nothing to that column sum. The same argument using now that $\pi_1 A \pi_2^{-1}$ is row-stochastic shows that C must in fact be an $(n-m) \times m$ block of zeros. Thus $\pi_1 A \pi_2^{-1} \in \mathbf{B}_{n-m} \times \mathbf{B}_m$ and so has a component count of at least 2. $\square$

Example 2.12. Consider the 8×8 doubly stochastic matrix:

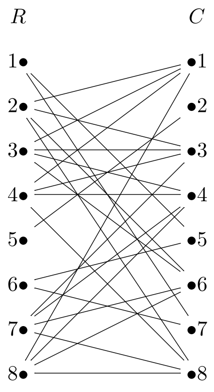
$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & .4 & 0 & .6 & 0 \\ .2 & 0 & .4 & 0 & 0 & .1 & 0 & .3 \\ .4 & 0 & .3 & .1 & 0 & .2 & 0 & 0 \\ .1 & 0 & .2 & .3 & 0 & 0 & 0 & .4 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .6 & 0 & .4 & 0 \\ 0 & 0 & .1 & .4 & 0 & .3 & 0 & .2 \\ .3 & 0 & 0 & .2 & 0 & .4 & 0 & .1 \end{pmatrix}.$$

Algorithm 2.1 of the preceding section yields a component count of $m = 3$. The members of the column-index partition $\mathcal{C}$, row-index partition $\mathcal{R}$, and component size partition $\mathfrak{n}$ of the integer $n = 8$, for the matrix A are displayed in the following table:

i	1	2	3
C_i	$\{1, 3, 4, 6, 8\}$	$\{2\}$	$\{5, 7\}$
R_i	$\{2, 3, 4, 7, 8\}$	$\{5\}$	$\{1, 6\}$
n_i	5	1	2

(2.2)

The weight graph of this A is:

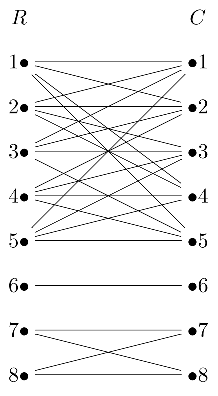


A is permutation equivalent to the block diagonal matrix:

$$B = \pi_{\mathcal{R}} A \pi_{\mathcal{C}}^{-1} = \begin{pmatrix} .2 & .4 & 0 & .1 & .3 & 0 & 0 & 0 \\ .4 & .3 & .1 & .2 & 0 & 0 & 0 & 0 \\ .1 & .2 & .3 & 0 & .4 & 0 & 0 & 0 \\ 0 & .1 & .4 & .3 & .2 & 0 & 0 & 0 \\ .3 & 0 & .2 & .4 & .1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .4 & .6 \\ 0 & 0 & 0 & 0 & 0 & 0 & .6 & .4 \end{pmatrix} \in \mathbf{B}_n(\mathfrak{n}) \cong \mathbf{B}_5 \times \mathbf{B}_1 \times \mathbf{B}_2 \subset \mathbf{B}_8,$$

where $\pi_{\mathcal{C}}$ and $\pi_{\mathcal{R}}$ can be read off as the permutations which send the second and third rows, respectively, of the above table (2.2) to $(1, \ldots, 8)$ (or, in cycle notation, $\pi_{\mathcal{C}} = (3\,2\,6\,4)\,(5\,7\,8)$ and $\pi_{\mathcal{R}} = (2\,1\,7\,4\,3)\,(5\,6\,8)$). Notice that the blocks of B are fully indecomposable.

Note that the weight graph of B is isomorphic to that of A but shows the components much more clearly:



3. Stochastic averaging does not decrease entropy

While for the most part entropy increases under stochastic averaging, it never decreases:

Proposition 3.1 (Shannon) [1]**.** *Let $A \in \mathbf{B}_n$ and $P \in \mathbf{\Sigma}_n$. Then the entropy of the stochastic average AP is not less than the entropy of P*: $H(AP) \geqslant H(P)$.

Proof. Say $P = (p_1, \ldots, p_n)^t$. The entropy of P can also be expressed as $H(P) = -\sum \varphi(p_i)$, where $\varphi(x) = x \log_2(x)$, extended (by continuity) so that $\varphi(0) = 0$. But since this $\varphi(x)$ is convex on the interval [0, 1] and A is column-stochastic,

$$\varphi\left(\sum_{j=1}^{n} a_{ij} p_j\right) \leqslant \sum_{j=1}^{n} a_{ij} \varphi(p_j) \tag{3.1}$$

holds for any i. We then have

$$\begin{aligned} H(AP) &= -\sum_{i=1}^{n} \varphi\left(\sum_{j=1}^{n} a_{ij} p_j\right), \\ &\geqslant -\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\varphi(p_j), \\ &= -\sum_{j=1}^{n}\left(\sum_{i=1}^{n} a_{ij}\right)\varphi(p_j), \\ &= -\sum_{j=1}^{n}\varphi(p_j), \\ &= H(P). \quad \square \end{aligned} \tag{3.2}$$

4. Three characterizations of Γ_n

4.1. ds-Matrices that respect probability distributions

Recall that given $\mathcal{J}, \mathcal{J}' \in \mathbf{\Phi}_n$, we say that $\mathcal{J}$ *refines* $\mathcal{J}'$ if $\forall J \in \mathcal{J}\ \exists J' \in \mathcal{J}'$ such that $J \subseteq J'$. In such a case, we write $\mathcal{J} \leqslant \mathcal{J}'$, while we write $\mathcal{J} < \mathcal{J}'$ if $\mathcal{J} \leqslant \mathcal{J}'$ and $\mathcal{J} \neq \mathcal{J}'$. $\mathbf{\Phi}_n$ becomes a partially ordered set with this ordering.

The following notion is crucial to the rest of this paper:

Definition 4.1. Let $A \in \mathbf{B}_n$ and $P \in \mathbf{\Sigma}_n$. We say that A *respects* P if the column-index partition $\mathcal{C}_A$ of A refines the coincidence partition $\mathcal{I}_P$ of P : $\mathcal{C}_A \leqslant \mathcal{I}_P$. We write

$$\mathscr{R}_n = \{(A, P) \in \mathbf{B}_n \times \mathbf{\Sigma}_n | \text{ A respects } P\}$$

for the set of respectful pairs.

Recall that Proposition 1.3 states that the action of $\mathbf{S}_n \times \mathbf{S}_n$ on $\mathbf{B}_n \times \mathbf{\Sigma}_n$ leaves $\mathbf{\Gamma}_n$ invariant. We now show:

Proposition 4.2. *$\mathscr{R}_n$ is invariant under the action of $\mathbf{S}_n \times \mathbf{S}_n$ on $\mathbf{B}_n \times \mathbf{\Sigma}_n$.*

Proof. We must show that for any $\pi_1, \pi_2 \in \mathbf{S}_n$, $A \in \mathbf{B}_n$, and $P \in \mathbf{\Sigma}_n$, A respects P if and only if $B = \pi_1 A \pi_2^{-1}$ respects $Q = \pi_2 P$. But for any permutation π_2, $\mathcal{C}_A$ refines $\mathcal{I}_P$ if and only if $\pi_2(\mathcal{C}_A)$ refines $\pi_2(\mathcal{I}_P)$. By Proposition 2.10, $\mathcal{C}_B = \pi_2(\mathcal{C}_A)$, while $\mathcal{I}_Q = \pi_2(\mathcal{I}_P)$, by the note following Definition 2.5. Hence $\mathcal{C}_A$ refines $\mathcal{I}_P$ if and only if $\mathcal{C}_B$ refines $\mathcal{I}_Q$. $\square$

One of our main goals is to show that $\mathbf{\Gamma}_n = \mathscr{R}_n$. First we prove that every $\mathbf{S}_n \times \mathbf{S}_n$-orbit in $\mathscr{R}_n$ intersects $\mathbf{B}_n(\mathfrak{n}) \times \mathbf{\Sigma}_n(\mathfrak{n})$, for some $\mathfrak{n} \in \phi_n$:

Proposition 4.3. *Suppose $(A, P) \in \mathscr{R}_n$. Let $\mathcal{I} = \{I_1, \ldots, I_l\} \in \mathbf{\Phi}_{n,l}$ denote the coincidence partition of P, with some order, and let $\mathfrak{n} \in \phi_{n,l}$ be the ordered integer partition induced by $\mathcal{I}$. Then there exists $\pi \in \mathbf{S}_n$ such that*

$$(\pi, \pi_{\mathcal{I}}) \cdot (A, P) \in \mathbf{B}_n(\mathfrak{n}) \times \mathbf{\Sigma}_n(\mathfrak{n}).$$

Proof. That $\pi_{\mathcal{I}} P \in \mathbf{\Sigma}_n(\mathfrak{n})$ follows from the Definition 2.3 of an unscrambling permutation. Now let $\mathcal{R}$ and $\mathcal{C}$ denote the row-index and column-index partitions of A. Define an ordered partition $\mathcal{J} = \{J_1, \ldots, J_l\}$ of $\{1, \ldots, n\}$ as follows: for $i \in \{1, \ldots, l\}$, let

$$J_i = \bigcup_{j \ni C_j \subset I_i} R_j.$$

Letting $\pi = \pi_{\mathcal{J}}$, we have $\pi A \pi_{\mathcal{I}}^{-1} \in \mathbf{B}_n(\mathfrak{n})$. $\square$

Note: Diagonal blocks of $\pi A \pi_{\mathcal{I}}^{-1} \in \mathbf{B}_n(\mathfrak{n})$ may not be fully indecomposable.

4.2. Main result – conditions for entropy-preserving stochastic averages

We shall need Theorem 90 in Hardy, Littlewood, and Polya's *Inequalities* [5, p. 74], specialized to the continuous, non-linear, convex function $\varphi(x) = x \log_2 x$ of Proposition 3.1.

Lemma 4.4. *Let $\{q_1, \ldots, q_k\}$ be a set of k positive weights that sum to one and let $\{p_1, \ldots, p_k\}$ be a set of k real numbers in the unit interval $[0, 1]$. Then*

$$\varphi\left(\sum_{j=1}^{k} q_j p_j\right) = \sum_{j=1}^{k} q_j \varphi(p_j)$$

if and only if the p_js are all equal.

Now we are in the position to state and prove our main result:

Theorem 4.5. *Let $A \in \mathbf{B}_n$ be an $n \times n$ doubly stochastic matrix and $P \in \mathbf{\Sigma}_n$ an n-long probability vector. The following statements are equivalent*[1]:

1. $H(AP) = H(P)$ *i.e.*, $(A, P) \in \mathbf{\Gamma}_n$
2. *A respects P* *i.e.*, $(A, P) \in \mathscr{R}_n$
3. $A^t AP = P$
4. $AP = \pi P$ *for some permutation matrix* $\pi \in \mathbf{S}_n$.

Proof. Let $\mathcal{C} = \{C_1, \ldots, C_m\}$ and $\mathcal{R} = \{R_1, \ldots, R_m\}$ be the column-index and row-index partitions induced by A, respectively, and let $\mathcal{I} = \{I_1, \ldots, I_l\}$ be the coincidence partition induced by P.

$\boxed{(1)\Rightarrow(2):}$ If $H(AP) = H(P)$, then inequality (3.2) is actually an equality. In particular, the first two lines of (3.2) yield

$$\sum_{i=1}^{n} \varphi\left(\sum_{j=1}^{n} a_{ij}p_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\varphi(p_j).$$

By the inequality (3.1), the summands here all change in the same direction; since the sums are equal and it cannot happen that one term increases while the other decreases, in fact the summands must all individually be equal. Since $\varphi(0) = 0$, we may as well restrict our attention only to those terms which are non-zero, *i.e.,*

$$\varphi\left(\sum_{\substack{j\ni\\ a_{ij}\neq 0}} a_{ij}p_j\right) = \sum_{\substack{j\ni\\ a_{ij}\neq 0}} a_{ij}\varphi(p_j).$$

But applying Lemma 4.4, we can conclude that the non-zero entries in the ith row of A must form a set of indices for which the p_j are all equal – that is, the non-zero indices in any row of A are a subset of some $I_j \in \mathcal{I}$.

Refer now to lines 7–10 of the Algorithm 2.1 constructing the components $\{W_1, \ldots, W_m\}$ of the weight graph W. The sets adjoined to some C_j each time through this loop are unions of sets of the form

$$C(r) = \{c \in C | a_{rc} > 0\}$$

for various values of r. We have just seen that such sets $C(r)$ are always each contained in a corresponding element of the partition $\mathcal{I}$, but it remains to see that the $C(r)$ whose union forms a C_j overlap sufficiently to ensure that they are subsets of *the same* element of $\mathcal{I}$.

Say some $c \in C_j$ and then that $r, r' \in R$ give edges (r, c) and (r', c) of W – in fact, the corresponding $C(r)$ and $C(r')$ are both to be adjoined to C_j. But the existence of these edges tells us that both rows r and r' of A have non-zero entries in column c (at least), *i.e.,* $c \in C(r) \cap C(r')$. This means that the constant value P takes on for $C(r)$ is the same as that value it takes on for $C(r')$, so $C(r) \cup C(r') \subseteq I_{j'} \in \mathcal{I}$ for some j'. Then continuing this way, also with new c' coming from the $C(r) \cup C(r')$ adjoined to C_j, builds all of C_j, which hence entirely lies in $I_{j'}$. Thus A respects P.

$\boxed{(2) \Rightarrow (3)\ \&\ (4):}$ Let π, $\pi_{\mathcal{I}}$, and $\mathfrak{n}$ be as in Proposition 4.3, so that $(B, Q) = (\pi A \pi_{\mathcal{I}}^{-1}, \pi_{\mathcal{I}}P) \in \mathscr{R}_n \bigcap (\mathbf{B}_n(\mathfrak{n}) \times \mathbf{\Sigma}_n(\mathfrak{n}))$. But then the first $n_1 \times n_1$ block of B is itself a ds-matrix, while Q restricted to $\{1, \ldots, n_1\}$ is a multiple of the uniform distribution, so this block leaves that part of Q unchanged. The same argument applies to the remaining blocks of B and Q, so in fact $BQ = Q$; exactly the same reasoning shows also that $B^tQ = Q$. But then

$$A^tAP = ({\pi_{\mathcal{I}}}^t B^t \pi)(\pi^t B \pi_{\mathcal{I}}){\pi_{\mathcal{I}}}^t Q = {\pi_{\mathcal{I}}}^t B^t BQ = {\pi_{\mathcal{I}}}^t B^t Q = {\pi_{\mathcal{I}}}^t Q = P$$

[1] The equivalence of (1) and (4) is Shannon's original result.

(using the fact that the transpose of a permutation matrix is its inverse), which is (3). Similarly,

$$AP = (\pi^t B\pi_{\mathcal{I}})\pi_{\mathcal{I}}{}^t Q = \pi^t BQ = \pi^t Q = (\pi^t \pi_{\mathcal{I}})P$$

which is (4), using the permutation $\pi^t\pi_{\mathcal{I}}$.

$\boxed{(3) \Rightarrow (1):}$ If $A^tAP = P$, then

$$H(P) = H(A^tAP) \geqslant H(AP) \geqslant H(P),$$

where the inequalities are two applications of Proposition 3.1, one using also the fact that transpose preserves $\mathbf{B}_n$.

$\boxed{(4) \Rightarrow (1):}$ If $AP = \pi P$ for some permutation matrix $\pi \in \mathbf{S}_n$, then $H(AP) = H(\pi P) = H(P)$. $\square$

5. Subsets of $\boldsymbol{\Gamma}_n$ of the form $(A, \cdot)$

Let us consider fibers of the projection map $\boldsymbol{\Gamma}_n \twoheadrightarrow \mathbf{B}_n$:

Definition 5.1. We write $\boldsymbol{\sigma}_A$ for the set of distributions $P \in \boldsymbol{\Sigma}_n$ whose stochastic averages with the fixed ds-matrix $A \in \mathbf{B}_n$ preserve the entropy of P.

Theorem 4.5 tells us that $\boldsymbol{\sigma}_A$ consists of the distributions respected by A, and is therefore completely determined by the column-index partition $\mathcal{C}_A$. In order to establish another description of $\boldsymbol{\sigma}_A$, we build a map as follows: Let $\mathcal{J} = \{J_1, \ldots, J_m\} \in \boldsymbol{\Phi}_{n,m}$ be indexed in natural order and let $\boldsymbol{n} = \{n_1, \ldots, n_m\} \in \boldsymbol{\phi}_{n,m}$ be the induced partition of the integer n. Write

$$T_{\mathcal{J}} : \mathbb{R}^m \to \mathbb{R}^n : (x_1, \ldots, x_m) \mapsto \sum_{i=1}^{m} \frac{x_i}{n_i} \left(\sum_{j \in J_i} \mathbf{e}_j^n \right),$$

(where, again, the $\mathbf{e}_j^n$ are the standard unit vectors in $\mathbb{R}^n$).

Here are some basic properties of this map:

Proposition 5.2. *Let $\mathcal{J} \in \boldsymbol{\Phi}_{n,m}$ be indexed in natural order. Then $T_{\mathcal{J}}$ is a rank m linear map which sends probability distributions to probability distributions. In fact, for any ds-matrix $A \in B_n$ whose column-index partition is $\mathcal{C}_A$ and whose component count is m, $T_{\mathcal{C}_A}(\boldsymbol{\Sigma}_m) = \boldsymbol{\sigma}_A$.*

Proof. Let the n_i be as above. The normalization factors of $\frac{1}{n_i}$ in the definition of $T_{\mathcal{J}}$ ensure that it carries $\boldsymbol{\Sigma}_m$ into $\boldsymbol{\Sigma}_n$. Next, if $Q \in \boldsymbol{\Sigma}_m$, then A respects $P = T_{\mathcal{C}_A}(Q)$, so that $T_{\mathcal{C}_A}(\boldsymbol{\Sigma}_m) \subseteq \boldsymbol{\sigma}_A$. Conversely, if A respects a distribution P, then P takes on at most m distinct values, and is constant on each set of indices C_i, so $\boldsymbol{\sigma}_A \subseteq T_{\mathcal{C}_A}(\boldsymbol{\Sigma}_m)$. $\square$

This map allows us to determine the geometry of $\boldsymbol{\sigma}_A$:

Theorem 5.3. *Let $A \in \mathbf{B}_n$ be an $n \times n$ ds-matrix and let m be the component count of its weight graph. Then*

- *$\boldsymbol{\sigma}_A \subset \boldsymbol{\Sigma}_n$ is an $(m-1)$-dimensional simplex and*
- *$\boldsymbol{\sigma}_A$ is a sub-complex of the barycentric subdivision of $\boldsymbol{\Sigma}_n$.*

Proof. The standard unit vectors $\{\mathbf{e}_j^m | j = 1, \ldots, m\}$ of $\mathbb{R}^m$, which are the vertices of $\boldsymbol{\Sigma}_m$, are each mapped by the nonsingular map $T_{\mathcal{C}_A}$ to the linearly independent vectors

$$\left\{T_{\mathcal{C}_A}(\mathbf{e}_i^m) = \frac{1}{n_i}\left(\sum_{j\in C_i}\mathbf{e}_j^n\right) \middle| i = 1, \ldots, m\right\} \subset \mathbb{R}^n.$$

These vectors are then the vertices of $\boldsymbol{\sigma}_A$, which therefore has the stated dimension. Furthermore, the ith of these vertices is the barycenter of the $(n_i - 1)$-dimensional face of $\boldsymbol{\Sigma}_n$ with vertices $\{\mathbf{e}_j \mid j \in C_i\}$, which yields the second statement of this theorem. □

Example 5.4. For the ds-matrix A from Example 2.12, the barycenters of the sub-simplices of $\boldsymbol{\Sigma}_8$ determined by $\mathcal{C}_A = \{C_1, C_2, C_3\} = \{\{1, 3, 4, 6, 8\}, \{2\}, \{5, 7\}\}$ are, respectively, the distributions:

$$\begin{aligned}
T_{\mathcal{C}_A}(\mathbf{e}_1^3) &= (1/5, \quad 0, \quad 1/5, \quad 1/5, \quad 0, \quad 1/5, \quad 0, \quad 1/5)^t\\
T_{\mathcal{C}_A}(\mathbf{e}_2^3) &= (0, \quad 1, \quad 0, \quad 0, \quad 0, \quad 0, \quad 0, \quad 0)^t\\
T_{\mathcal{C}_A}(\mathbf{e}_3^3) &= (0, \quad 0, \quad 0, \quad 0, \quad 1/2, \quad 0, \quad 1/2, \quad 0)^t.
\end{aligned}$$

The convex hull of these three distributions is $\boldsymbol{\sigma}_A$. In particular, let $Q = (q_1, q_2, q_3) \in \boldsymbol{\Sigma}_3$ be any distribution and write $P = T_{\mathcal{C}_A}(Q) = \sum_{i=1}^3 q_i T_{\mathcal{C}_A}(\mathbf{e}_i^3)$. Then

$$P = (q_1/5, \quad q_2, \quad q_1/5, \quad q_1/5, \quad q_3/2, \quad q_1/5 \quad q_3/2 \quad q_1/5)^t$$

while

$$AP = (q_3/2, \quad q_1/5, \quad q_1/5, \quad q_1/5, \quad q_2, \quad q_3/2, \quad q_1/5, \quad q_1/5)^t.$$

This AP is a permutation of P, hence $H(AP) = H(P)$, and so $P \in \boldsymbol{\sigma}_A$.

Example 5.5. If $A \in \mathbf{S}_n$, then A has a component count of n and $\boldsymbol{\sigma}_A = \boldsymbol{\Sigma}_n$.

Example 5.6. If $A \in \mathbf{B}_n$ has a strictly positive row or column, then A has a component count of 1 and $\boldsymbol{\sigma}_A = \{U\}$.

Example 5.7. If A is a direct sum of m fully indecomposable ds-matrices, then A has a component count of m by Theorem 2.11 and $\boldsymbol{\sigma}_A$ is an $(m-1)$-simplex.

6. Subsets of $\boldsymbol{\Gamma}_n$ of the form $(\cdot, P)$

Now we study fibers of the projection $\boldsymbol{\Gamma}_n \twoheadrightarrow \boldsymbol{\Sigma}_n$:

Definition 6.1. We write $\mathbf{b}_P$ for the set of ds-matrices $A \in \mathbf{B}_n$ whose stochastic averages with the fixed distribution $P \in \boldsymbol{\Sigma}_n$ preserve its entropy.

Again, Theorem 4.5 says that $\mathbf{b}_P$ consists of the ds-matrices that respect P. Here is a concrete description of $\mathbf{b}_P$:

Theorem 6.2. *Let $P \in \boldsymbol{\Sigma}_n$ be a probability distribution. Choose an ordering of its coincidence partition $\mathcal{I} \in \boldsymbol{\Phi}_{n,l}$ and let $\mathfrak{n} = \{n_1, \ldots, n_l\} \in \phi_{n,l}$ be the induced integer partition. Let $\pi_{\mathcal{I}} \in \mathbf{S}_n$ be the corresponding unscrambling permutation. Then we have*

$$\mathbf{b}_P = \bigcup_{\pi\mathbf{S}_n(\mathfrak{n})\in\mathbf{S}_n/\mathbf{S}_n(\mathfrak{n})} \pi \cdot \mathbf{B}_n(\mathfrak{n}) \cdot \pi_{\mathcal{I}}, \tag{6.1}$$

which expresses $\mathbf{b}_P$ as a union of $n!/n_1!n_2!\cdots n_l!$ pairwise disjoint faces of the polytope $\mathbf{B}_n$, each of which is isomorphic to $\mathbf{B}_n(\mathfrak{n})$. Hence, also, $\mathbf{b}_P$ has dimension $\sum_{k=1}^l (n_k - 1)^2$.

Proof. Proposition 4.3 tells us that for every $A \in \mathbf{b}_P$, there is a $\pi \in \mathbf{S}_n$ such that $\pi A \pi_{\mathcal{I}}^{-1} \in \mathbf{B}_n(\mathfrak{n})$. In other words,

$$\mathbf{b}_P \subseteq \{\pi B \pi_{\mathcal{I}} | \quad \pi \in \mathbf{S}_n, B \in \mathbf{B}_n(\mathfrak{n})\} = \mathbf{S}_n \cdot \mathbf{B}_n(\mathfrak{n}) \cdot \pi_{\mathcal{I}}.$$

Since the opposite inclusion holds in light of Theorem 4.5 and the definition of respect, we actually have that $\mathbf{b}_P = \mathbf{S}_n \cdot \mathbf{B}_n(\mathfrak{n}) \cdot \pi_{\mathcal{I}}$. However, this description is not the most efficient, since the $\mathbf{S}_n$-action is not effective: all of the elements of $\mathbf{S}_n(\mathfrak{n})$ fix $\mathbf{B}_n(\mathfrak{n})$. Dividing by this subgroup produces the equality (6.1).

Since $\#(\mathbf{S}_n(\mathfrak{n})) = n_1!n_2! \cdots n_l!$ and $\dim(\mathbf{B}_n(\mathfrak{n})) = \sum_{k=1}^{l}(n_k - 1)^2$, we shall be done if we can show that the different faces $\pi \cdot \mathbf{B}_n(\mathfrak{n}) \cdot \pi_{\mathcal{I}}$ for representatives π of different cosets in $\mathbf{S}_n/\mathbf{S}_n(\mathfrak{n})$ are disjoint. For this, say $B_1, B_2 \in \mathbf{B}_n(\mathfrak{n})$ and $\pi_1, \pi_2 \in \mathbf{S}_n$ have $\pi_1 B_1 \pi_{\mathcal{I}} = \pi_2 B_2 \pi_{\mathcal{I}}$. But then $(\pi_2^{-1}\pi_1)B_1 = B_2$, which is a permutation matrix $\pi_2^{-1}\pi_1$ acting on the left on a matrix in $\mathbf{B}_n(\mathfrak{n})$ and yielding again a matrix in $\mathbf{B}_n(\mathfrak{n})$. Look at a row in B_1, in the block of rows defined by jth subset in the partition which defines $\mathbf{B}_n(\mathfrak{n})$. Since the entire block is a $n_j \times n_j$ ds-matrix, there must be at least one non-zero element in this row. But then if we permute the rows by left-multiplication with $\pi_2^{-1}\pi_1$ and have again an element of $\mathbf{B}_n(\mathfrak{n})$, the row in question must not have been moved out of the jth block. Therefore $\pi_2^{-1}\pi_1 \in \mathbf{S}_n(\mathfrak{n})$ and we are done. □

Example 6.3. If $P = (1, 0, \ldots, 0)^t \in \mathbf{\Sigma}_n$, then $l = 2$ and $\mathbf{b}_P$ is the disjoint union of the n faces $\pi(\mathbf{B}_1 \times \mathbf{B}_{n-1})$ of $\mathbf{B}_n$ as π runs over the set of the permutation matrices obtained by exchanging the first and ith rows of the identity matrix, for $i \in \{1, \ldots, n\}$.

Example 6.4. For the uniform distribution $U \in \mathbf{\Sigma}_n$, we have: $l = 1, \mathfrak{n} = \{n\}, \mathbf{B}_n(\mathfrak{n}) = \mathbf{B}_n, \mathbf{S}_n(\mathfrak{n}) = \mathbf{S}_n$, and the decomposition in (6.1) yields $\mathbf{b}_U$ as the one face $\mathbf{B}_n$.

Example 6.5. If $P \in \mathbf{\Sigma}_n$ takes n distinct values, then: $l = n, \mathfrak{n} = \{1, \ldots, 1\}, \mathbf{B}_n(\mathfrak{n}) = \{\mathrm{Id}_{n\times n}\} = \mathbf{S}_n(\mathfrak{n})$, and the decomposition in (6.1) writes $\mathbf{b}_P$ as the union of $n!$ 0-dimensional faces, $\mathbf{b}_P = \mathbf{S}_n \cdot \{\mathrm{Id}_{n\times n}\} = \mathbf{S}_n \subset \mathbf{B}_n$.

7. The topology of $\mathbf{\Gamma}_n$

It is natural to think of the topology and linear structure of $\mathbf{\Gamma}_n$ as coming from its realization as a subset of $\mathbf{B}_n \times \mathbf{\Sigma}_n \subset \mathbb{R}^{n^2+n}$. Then we can ask how the different fibers discussed in the last two sections fit together. The answer is quite simple:

Theorem 7.1. *$\mathbf{\Gamma}_n$ is PL-contractible.*[2]

Proof. Each $\boldsymbol{\sigma}_A$ contains U and is convex. Also, as we have seen in Example 6.4, $\mathbf{b}_U = \mathbf{B}_n$, which is also convex. So for any $B \in \mathbf{B}_n$, we can contract first each of the fibers of $\mathbf{\Gamma}_n \twoheadrightarrow \mathbf{B}_n$ to the distribution U, then contract the base of this fibration to the point B. Explicitly, define $G_B : \mathbf{B}_n \times \mathbf{\Sigma}_n \times [0, 1] \to \mathbf{B}_n \times \mathbf{\Sigma}_n$ by

$$G_B(A, P, t) = \begin{cases} (A, \ (1-2t)P + 2tU) & \text{if } 0 \leqslant t \leqslant 1/2 \\ ((2-2t)A + (2t-1)B, \ U) & \text{if } 1/2 \leqslant t \leqslant 1. \end{cases}$$

This $G_B|_{\mathbf{\Gamma}_n}$ is a PL-homotopy between the identity map on $\mathbf{\Gamma}_n$ and the constant map $\mathbf{\Gamma}_n \to \{(B, U)\}$. □

8. Enumerating the distinct fibers of the maps $\mathbf{B}_n \twoheadleftarrow \mathbf{\Gamma}_n \twoheadrightarrow \mathbf{\Sigma}_n$

Let us now examine all the fibers from Sections 5 and 6 separately. It is convenient to have notation for the sets of such fibers.

[2] See [6] for a classical background on the PL category.

Definition 8.1. We write

$$s_n = \{\boldsymbol{\sigma}_A | A \in \mathbf{B}_n\} \subset 2^{\boldsymbol{\Sigma}_n}$$
$$\beta_n = \{\mathbf{b}_P | P \in \boldsymbol{\Sigma}_n\} \subset 2^{\mathbf{B}_n}$$

and endow each with a partial ordering coming from set inclusion.

We have:

Theorem 8.2. *There exist bijections $F : \boldsymbol{\Phi}_n \to s_n$, which reverses order, and $G : \boldsymbol{\Phi}_n \to \beta_n$, which preserves order.*

Proof. First, given $\mathcal{J} \in \boldsymbol{\Phi}_n$, define a ds-matrix $f(\mathcal{J}) \in \mathbf{B}_n$ as follows: for $r, c \in \{1, \ldots, n\}$ let

$$f(\mathcal{J})_{rc} = \begin{cases} 1/\#(J) & \text{if } r \text{ and } c \text{ are in the same } J \in \mathcal{J} \\ 0 & \text{otherwise.} \end{cases}$$

In order to understand $f : \boldsymbol{\Phi}_n \to \mathbf{B}_n$, consider the case of a partition $\mathfrak{n} = \{n_1, \ldots, n_m\} \in \phi_{n,m}$ of n, with some chosen ordering, and then $\mathcal{N} \in \boldsymbol{\Phi}_{n,m}$, the set partition induced by $\mathfrak{n}$. Now $f(\mathcal{N})$ is a block diagonal matrix, with successive blocks being of size $n_i \times n_i$ and filled with the entry $1/n_i$. Therefore the weight graph of $f(\mathcal{N})$ has m components, each of which is the complete bipartite graph on $2n_i$ vertices and hence the column-index partition of $f(\mathcal{N})$ is $\mathcal{N}$.

Any other $\mathcal{J} \in \boldsymbol{\Phi}_n$ is a permutation of a partition $\mathcal{N}$ as just examined, so $f(\mathcal{J})$ is conjugate to a block diagonal matrix and still has an isomorphic weight graph. Thus $\mathcal{C}_f(\mathcal{J}) = \mathcal{J}$. This means that f is one-to-one since following it with the function that takes a ds-matrix to its column-index partition yields the identity on $\boldsymbol{\Phi}_n$.

Out of this f, build the desired function

$$F : \boldsymbol{\Phi}_n \to s_n : \mathcal{J} \mapsto \boldsymbol{\sigma}_{f(\mathcal{J})}.$$

Similarly, starting with $\mathcal{J} \in \boldsymbol{\Phi}_{n,m}$, define a distribution $g(\mathcal{J}) \in \boldsymbol{\Sigma}_n$ as follows: index $\mathcal{J}$ as $\{J_1, \ldots, J_m\}$ in natural order. Then, for $j \in \{1, \ldots, n\}$, set

$$g(\mathcal{J})_j = \frac{i}{\sum_{k=1}^m k\#(J_k)} \quad \text{if } j \in J_i.$$

The coincidence partition of $g(\mathcal{J})$ is $\mathcal{J}$, *i.e.*, $\mathcal{I}_{g(\mathcal{J})} = \mathcal{J}$. The function $g : \boldsymbol{\Phi}_n \to \boldsymbol{\Sigma}_n$ is one-to-one since following it with the function that takes a distribution into its coincidence partition is the identity on $\boldsymbol{\Phi}_n$. As above, we build

$$G : \boldsymbol{\Phi}_n \to \beta_n : \mathcal{J} \mapsto \mathbf{b}_{g(\mathcal{J})}$$

out of g.

Since, for any $A \in \mathbf{B}_n$, $\boldsymbol{\sigma}_A$ is completely determined by its column-index partition $\mathcal{C}_A$ and $\mathcal{C}_A = \mathcal{C}_{f(\mathcal{C}_A)}$ we have $F(\mathcal{C}_A) = \boldsymbol{\sigma}_{f(\mathcal{C}_A)} = \boldsymbol{\sigma}_A$ and so F is surjective. To see that it is injective, assume $F(\mathcal{J}) = F(\mathcal{J}')$. The distribution $g(\mathcal{J})$ is in $\boldsymbol{\sigma}_{f(\mathcal{J})}$ since the column-index partition of $f(\mathcal{J})$ refines (indeed, equals) the coincidence partition of $g(\mathcal{J})$. But $\boldsymbol{\sigma}_{f(\mathcal{J})} = \boldsymbol{\sigma}_{f(\mathcal{J}')}$, so it follows that $\mathcal{J}'$ refines $\mathcal{J}$. Symmetry then tells us that $\mathcal{J} = \mathcal{J}'$.

Likewise, for any $P \in \boldsymbol{\Sigma}_n$, $\mathbf{b}_P$ is completely determined by its coincidence partition $\mathcal{I}_P$ and $\mathcal{I}_P = \mathcal{I}_{g(\mathcal{I}_P)}$, so we have $G(\mathcal{I}_P) = \mathbf{b}_{g(\mathcal{I}_P)} = \mathbf{b}_P$ and thus G is also surjective. To see that G is one-to-one, let $G(\mathcal{J}) = G(\mathcal{J}')$. The ds-matrix $f(\mathcal{J})$ is in $\mathbf{b}_{g(\mathcal{J})}$ since its column-index partition refines the coincidence partition of $g(\mathcal{J})$. But $\mathbf{b}_{g(\mathcal{J})} = \mathbf{b}_{g(\mathcal{J}')}$, so it follows that $\mathcal{J}$ refines $\mathcal{J}'$. By symmetry again we have $\mathcal{J} = \mathcal{J}'$.

Now take $\mathcal{J}, \mathcal{J}' \in \boldsymbol{\Phi}_n$ satisfying $\mathcal{J} \leqslant \mathcal{J}'$. Since $F(\mathcal{J}') = \boldsymbol{\sigma}_{f(\mathcal{J}')}$ consists of distributions P for which $\mathcal{J}' = \mathcal{C}_{f(\mathcal{J}')} \leqslant \mathcal{I}_P$, it follows that $F(\mathcal{J}') \subseteq F(\mathcal{J})$. Similarly, $G(\mathcal{J}) = \mathbf{b}_{g(\mathcal{J})}$ consists of ds-matrices A for which $\mathcal{C}_A \leqslant \mathcal{I}_{g(\mathcal{J})} = \mathcal{J}$, so $G(\mathcal{J}) \subseteq G(\mathcal{J}')$. $\square$

9. Partitioning $\boldsymbol{\Gamma}_n$ into finitely many product subsets

We defined in Sections 5 and 6 fibers of maps from $\boldsymbol{\Gamma}_n$ onto $\mathbf{B}_n$ and $\boldsymbol{\Sigma}_n$, out of which we can write $\boldsymbol{\Gamma}_n$ as an uncountable union of product subsets in the two ways

$$\boldsymbol{\Gamma}_n = \bigcup_{A \in \mathbf{B}_n} \{A\} \times \boldsymbol{\sigma}_A = \bigcup_{P \in \boldsymbol{\Sigma}_n} \mathbf{b}_P \times \{P\}.$$

These descriptions are highly inefficient, however, since many points of $\boldsymbol{\Gamma}_n$ appear many times. We are in fact able to show the:

Theorem 9.1. *Each of the following expressions decomposes $\boldsymbol{\Gamma}_n \subset \mathbf{B}_n \times \boldsymbol{\Sigma}_n$ into a finite number of disjoint product subsets*:

$$\boldsymbol{\Gamma}_n = \bigcup_{\mathcal{J} \in \boldsymbol{\Phi}_n} \mathbf{b}_{g(\mathcal{J})} \times \left(\boldsymbol{\sigma}_{f(\mathcal{J})} \Big\backslash \bigcup_{\substack{\mathcal{J}' \in \boldsymbol{\Phi}_n \\ \ni \mathcal{J} < \mathcal{J}'}} \boldsymbol{\sigma}_{f(\mathcal{J}')} \right) = \bigcup_{\mathcal{J} \in \boldsymbol{\Phi}_n} \left(\mathbf{b}_{g(\mathcal{J})} \Big\backslash \bigcup_{\substack{\mathcal{J}' \in \boldsymbol{\Phi}_n \\ \ni \mathcal{J}' < \mathcal{J}}} \mathbf{b}_{g(\mathcal{J}')} \right) \times \boldsymbol{\sigma}_{f(\mathcal{J})},$$

where f and g are as defined in the previous section.

Proof. Consider the function

$$F_0 : \boldsymbol{\Phi}_n \to 2^{\boldsymbol{\Sigma}_n} : \mathcal{J} \mapsto \{P \in \boldsymbol{\Sigma}_n \mid \mathcal{I}_P = \mathcal{J}\} \ .$$

A distribution in $\boldsymbol{\Sigma}_n$ has only one coincidence partition, so F_0 carries distinct $\mathcal{J}$ into non-intersecting subsets. Also, $P \in F_0(\mathcal{I}_P)$, so $\{F_0(\mathcal{J}) \mid \mathcal{J} \in \boldsymbol{\Phi}_n\}$ is a finite decomposition of $\boldsymbol{\Sigma}_n$. Consider also the function

$$G_0 : \boldsymbol{\Phi}_n \to 2^{\mathbf{B}_n} : \mathcal{J} \mapsto \{A \in \mathbf{B}_n \mid \mathcal{C}_A = \mathcal{J}\} \ .$$

Reasoning as above, we have that $\{G_0(\mathcal{J}) \mid \mathcal{J} \in \boldsymbol{\Phi}_n\}$ is a finite decomposition of $\mathbf{B}_n$.

By Theorem 4.5, $\boldsymbol{\Gamma}_n$ is the union over $\mathcal{J} \in \boldsymbol{\Phi}_n$ of the disjoint sets

$$\{(A, P) \in \mathbf{B}_n \times \boldsymbol{\Sigma}_n \mid \mathcal{C}_A \text{ refines } \mathcal{I}_P \text{ and } \mathcal{I}_P = \mathcal{J}\}$$

i.e., the sets

$$\{A \in \mathbf{B}_n \mid \mathcal{C}_A \text{ refines } \mathcal{J}\} \times \{P \in \boldsymbol{\Sigma}_n \mid \mathcal{I}_P = \mathcal{J}\} = \mathbf{b}_{g(\mathcal{J})} \times F_0(\mathcal{J}).$$

Similarly, $\boldsymbol{\Gamma}_n$ is also the union over $\mathcal{J} \in \boldsymbol{\Phi}_n$ of the disjoint sets

$$\{(A, P) \in \mathbf{B}_n \times \boldsymbol{\Sigma}_n \mid \mathcal{C}_A \text{ refines } \mathcal{I}_P \text{ and } \mathcal{C}_A = \mathcal{J}\}$$

i.e., the sets

$$\{A \in \mathbf{B}_n \mid \mathcal{C}_A = \mathcal{J}\} \times \{P \in \boldsymbol{\Sigma}_n \mid \mathcal{J} \text{ refines } \mathcal{I}_P\} = G_0(\mathcal{J}) \times \boldsymbol{\sigma}_{f(\mathcal{J})}.$$

Say that we have a pair $\mathcal{J}, \mathcal{J}' \in \boldsymbol{\Phi}_n$ for which $\mathcal{J} < \mathcal{J}'$. If $P \in \boldsymbol{\sigma}_{f(\mathcal{J}')}$ then $P \in \boldsymbol{\sigma}_{f(\mathcal{J})}$. Furthermore $g(\mathcal{J})$ is in $\boldsymbol{\sigma}_{f(\mathcal{J})}$ but not in $\boldsymbol{\sigma}_{f(\mathcal{J}')}$, so $\boldsymbol{\sigma}_{f(\mathcal{J}')} \subsetneq \boldsymbol{\sigma}_{f(\mathcal{J})}$. Similarly, if $A \in \mathbf{b}_{g(\mathcal{J})}$ then $A \in \mathbf{b}_{g(\mathcal{J}')}$ and $f(\mathcal{J})$ is in $\mathbf{b}_{g(\mathcal{J}')}$ but not in $\mathbf{b}_{g(\mathcal{J})}$. Hence $\mathbf{b}_{g(\mathcal{J})} \subsetneq \mathbf{b}_{g(\mathcal{J}')}$.

In particular, for any $\mathcal{J} \in \boldsymbol{\Phi}_n$

$$F_0(\mathcal{J}) = \boldsymbol{\sigma}_{f(\mathcal{J})} \Big\backslash \bigcup_{\mathcal{J}' \in \boldsymbol{\Phi}_n \ni \mathcal{J} < \mathcal{J}'} \boldsymbol{\sigma}_{f(\mathcal{J}')}$$

and

$$G_0(\mathcal{J}) = \mathbf{b}_{g(\mathcal{J})} \Big\backslash \bigcup_{\mathcal{J}' \in \mathbf{\Phi}_n \ni \ \mathcal{J}' < \mathcal{J}} \mathbf{b}_{g(\mathcal{J}')}$$

from which follow the claimed finite decompositions of $\mathbf{\Gamma}_n$. $\square$

Example 9.2. Let us construct the two decompositions of $\mathbf{\Gamma}_2$.

There are only two partitions of {1,2}. They, and their images under f and g, are:

$$\begin{array}{llcl} \mathbf{\Phi}_2: & \{\{1\},\{2\}\} & < & \{\{1,2\}\} \\ \\ f: & \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & & \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \\ \\ g: & (1/3, 2/3)^t & & (1/2, 1/2)^t [= U]. \end{array}$$

Therefore, putting the terms in the unions of Theorem 9.1 in increasing order with regards to the refinement ordering " $<$ " of partitions,

$$\mathbf{\Gamma}_2 = [\mathbf{S}_2 \times (\mathbf{\Sigma}_2 \backslash \{U\})] \cup [\mathbf{B}_2 \times \{U\}] = [\mathbf{S}_2 \times \mathbf{\Sigma}_2] \cup [(\mathbf{B}_2 \backslash \mathbf{S}_2) \times \{U\}], \tag{9.1}$$

using (repeatedly) the examples in Sections 5 and 6. Since $\mathbf{\Gamma}_2 \subset \mathbf{B}_2 \times \mathbf{\Sigma}_2 \subset \mathbb{R}^2$, it is easy to make a picture of these decompositions: see Fig. 1.

Example 9.3. Now let us consider the case of $\mathbf{\Gamma}_3$. It is convenient to use some notation as follows: write $\mathbf{S}_3 = \mathbf{G} \cup \mathbf{H}$, where $\mathbf{G}$ denotes the cyclic subgroup of order 3 with elements

$$id = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \rho = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \rho^2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

and $\mathbf{H}$ denotes the set of order 2 elements of $\mathbf{S}_3$, being

$$\tau_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \tau_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \text{and} \quad \tau_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

To construct the two decompositions of $\mathbf{\Gamma}_3$ given by Theorem 9.1, it suffices to compute $\boldsymbol{\sigma}_{f(\mathcal{J})}$ and $\mathbf{b}_{g(\mathcal{J})}$ for each of the five partitions $\mathcal{J} \in \mathbf{\Phi}_3$. First, here are the partitions and their images under the maps f and g from Section 8:

$$\begin{array}{llclcl} \mathbf{\Phi}_3: & \{1\},\{2\},\{3\} & < & \begin{array}{c} \{1\},\{2,3\} \\ \{1,3\},\{2\} \\ \{1,2\},\{3\} \end{array} & < & \{1,2,3\} \\ \\ f: & id & & \begin{array}{c} 1/2(id+\tau_1) \\ 1/2(id+\tau_2) \\ 1/2(id+\tau_3) \end{array} & & 1/3(id+\rho+\rho^2) \\ \\ g: & (1/6,\ 1/3,\ 1/2)^t & & \begin{array}{c} (1/5,\ 2/5,\ 2/5)^t \\ (1/4,\ 1/2,\ 1/4)^t \\ (1/4,\ 1/4,\ 1/2)^t \end{array} & & (1/3,\ 1/3,\ 1/3)^t = U \end{array} \tag{9.2}$$

According to the examples and calculations in Section 5, we can compute $\boldsymbol{\sigma}_{f(\cdot)}$ for the above partitions:

$$\begin{array}{lll} \boldsymbol{\sigma}_{f(\cdot)}: \quad \mathbf{\Sigma}_3 & \begin{array}{c} \mathbf{e}_1 \star 1/2(\mathbf{e}_2+\mathbf{e}_3) \\ \mathbf{e}_2 \star 1/2(\mathbf{e}_1+\mathbf{e}_3) \\ \mathbf{e}_3 \star 1/2(\mathbf{e}_1+\mathbf{e}_2) \end{array} & \{U\} \end{array}$$

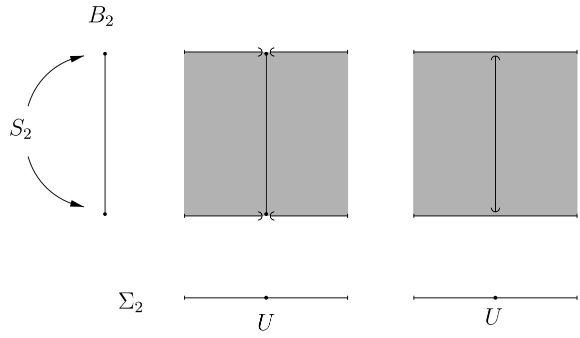


Fig. 1. Decompositions of $\boldsymbol{\Gamma}_2$: the two sideways "H"s are the two versions of $\boldsymbol{\Gamma}_2$ as given in (9.1), shown as sitting inside the square $\mathbf{B}_2 \times \boldsymbol{\Sigma}_2$.

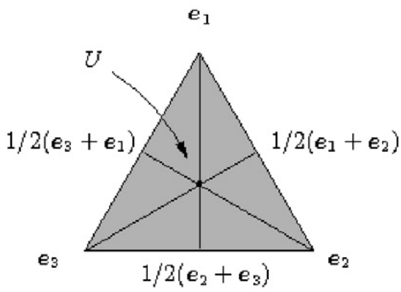


Fig. 2. The probability simplex $\boldsymbol{\Sigma}_3$, with the possible values of $\boldsymbol{\sigma}_{f(\cdot)}$ shown.

(here we suppress the superscript 3 on the basis vectors) where the join ($\star$) of a pair of points yields the line segment connecting them. Geometrically, these values of $\boldsymbol{\sigma}_{f(\cdot)}$ are: all of $\boldsymbol{\Sigma}_3$ – which is an equilateral triangle; that triangle's three medians; and the centroid of the triangle; see Fig. 2.

Turning to the $\mathbf{b}_{g(\cdot)}$, we shall apply Theorem 6.2, for which we need first to record the unscrambling permutations $\pi_{\mathcal{I}}$, again corresponding to the partitions in $\boldsymbol{\Phi}_3$, each with the ordering assigned by (9.2):

$$\begin{array}{llll} \pi_{\mathcal{I}}: & id & \begin{array}{l} id \\ \tau_1 \\ id \end{array} & id \end{array}$$

Note next that the partitions of 3 induced by the partitions $\mathcal{J} \in \boldsymbol{\Phi}_3$ are, reading across the $\boldsymbol{\Phi}_3$ row above:

$$\begin{array}{llll} \mathfrak{n}: & \{1,1,1\} & \begin{array}{l} \{1,2\} \\ \{2,1\} \\ \{2,1\} \end{array} & \{3\} \end{array}$$

These partitions of 3 yield corresponding adapted permutations and ds-matrices (see Definition 2.4):

$$\begin{array}{llll} \mathbf{S}_3(\mathfrak{n}): & \{id\} & \begin{array}{l} \{id,\ \tau_1\} \\ \{id,\ \tau_3\} \\ \{id,\ \tau_3\} \end{array} & \mathbf{S}_3 \end{array}$$

$$\begin{array}{llll} \mathbf{B}_3(\mathfrak{n}): & \{id\} & \begin{array}{l} id \star \tau_1 \\ id \star \tau_3 \\ id \star \tau_3 \end{array} & \mathbf{B}_3 \end{array}$$

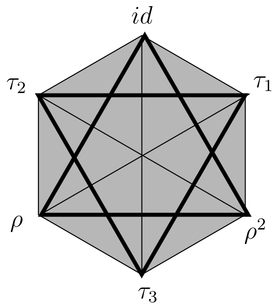


Fig. 3. A planar representation of the four-dimensional Birkhoff polytope $\mathbf{B}_3$: the six thick lines, the three thin lines and the six edges of the hexagon correspond to the 15 edges of $\mathbf{B}_3$. In four dimensions these lines intersect only at the six vertices.

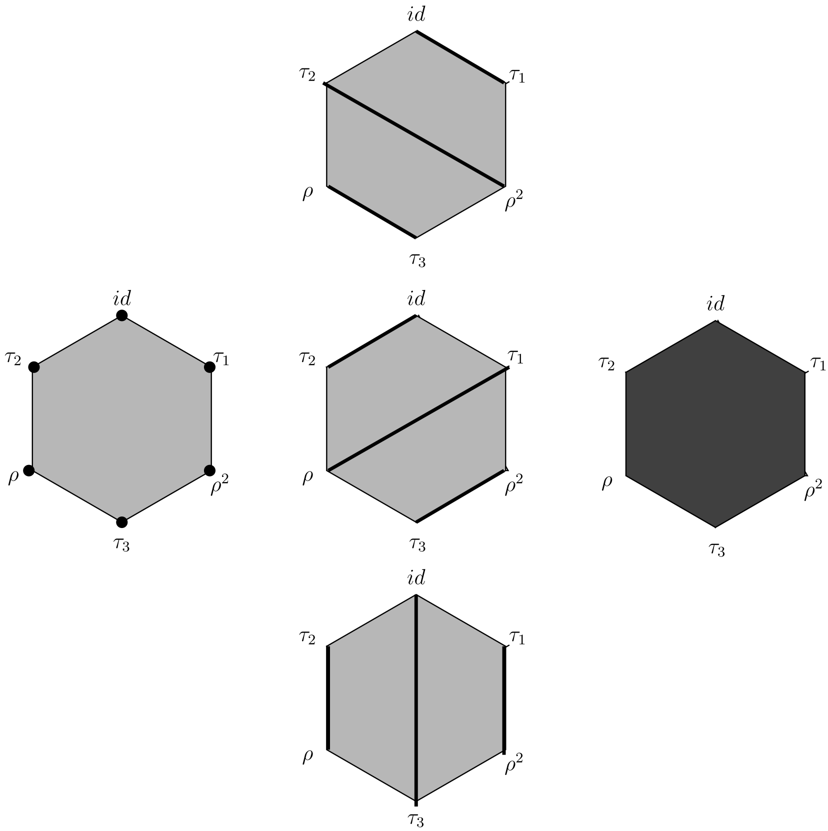


Fig. 4. The subsets $\mathbf{b}_{g(\cdot)}$ of $\mathbf{B}_3$ laid out corresponding to the partitions in Eq. (9.2)

The last ingredient we need is a complete set of representatives of the coset spaces:

$$\mathbf{S}_3/\mathbf{S}_3(\mathfrak{n}): \quad \mathbf{S}_3 \qquad \begin{array}{c}\{id,\ \tau_2,\ \tau_3\}\\ \{id,\ \tau_1,\ \tau_2\}\\ \{id,\ \tau_1,\ \tau_2\}\end{array} \qquad \{id\}$$

Putting all this information together into Theorem 6.2, we find:

$$\mathbf{b}_{g(\cdot)}: \quad \mathbf{S}_3 \qquad \begin{matrix} (id \star \tau_1) \cup (\rho \star \tau_3) \cup (\rho^2 \star \tau_2) \\ (id \star \tau_2) \cup (\rho \star \tau_1) \cup (\rho^2 \star \tau_3) \\ (id \star \tau_3) \cup (\rho \star \tau_2) \cup (\rho^2 \star \tau_1) \end{matrix} \qquad \mathbf{B}_3$$

In order to visualize how each $\mathbf{b}_{g(\cdot)}$ sits in $\mathbf{B}_3$, we need a planar representation of this four-dimensional polytope. Following [7], Fig. 3 is a display of the 6 vertexes and 15 edges of $\mathbf{B}_3$. There are 9 short edges and 6 long edges (drawn here with thick lines). The long edges enclose two triangles within $\mathbf{B}_3$. One of these triangles is the convex hull of the three permutations that constitute $\mathbf{G}$, the order three cyclic subgroup of $\mathbf{S}_3$. The other triangle is the convex hull of $\mathbf{H}$, the set of order two elements of $\mathbf{S}_3$. These two planes are orthogonal subsets of $\mathbf{B}_3$, intersecting at a single point in $\mathbf{B}_3$ – the ds-matrix whose entries are all equal to $\frac{1}{3}$.

Based on this planar representation, Fig. 4 shows the five sets of the form $\mathbf{b}_{g(\cdot)}$ in $\mathbf{B}_3$. They are arranged so as to reflect the location of the five partitions in the same manner as they are displayed on the page in Eq. (9.2).

Theorem 9.1 gives us the following two decompositions of $\mathbf{\Gamma}_3$:

$$\begin{aligned} \mathbf{\Gamma}_3 = \quad & \mathbf{S}_3 \times \Big(\mathbf{\Sigma}_3 \setminus [(\mathbf{e}_1 \star 1/2(\mathbf{e}_2+\mathbf{e}_3)) \cup (\mathbf{e}_2 \star 1/2(\mathbf{e}_1+\mathbf{e}_3)) \cup (\mathbf{e}_3 \star 1/2(\mathbf{e}_1+\mathbf{e}_2))]\Big) \\ & \cup \Big[(id \star \tau_1) \cup (\rho \star \tau_3) \cup (\rho^2 \star \tau_2)\Big] \times \Big((\mathbf{e}_1 \star 1/2(\mathbf{e}_2+\mathbf{e}_3)) \setminus \{U\}\Big) \\ & \cup \Big[(id \star \tau_2) \cup (\rho \star \tau_1) \cup (\rho^2 \star \tau_3)\Big] \times \Big((\mathbf{e}_2 \star 1/2(\mathbf{e}_1+\mathbf{e}_3)) \setminus \{U\}\Big) \\ & \cup \Big[(id \star \tau_3) \cup (\rho \star \tau_2) \cup (\rho^2 \star \tau_1)\Big] \times \Big((\mathbf{e}_3 \star 1/2(\mathbf{e}_1+\mathbf{e}_2)) \setminus \{U\}\Big) \\ & \cup \mathbf{B}_3 \times \{U\} \end{aligned}$$

and

$$\begin{aligned} \mathbf{\Gamma}_3 = \quad & \mathbf{S}_3 \times \mathbf{\Sigma}_3 \\ & \cup \Big[\Big((id \star \tau_1) \cup (\rho \star \tau_3) \cup (\rho^2 \star \tau_2)\Big) \setminus \mathbf{S}_3\Big] \times \Big(\mathbf{e}_1 \star 1/2(\mathbf{e}_2+\mathbf{e}_3)\Big) \\ & \cup \Big[\Big((id \star \tau_2) \cup (\rho \star \tau_1) \cup (\rho^2 \star \tau_3)\Big) \setminus \mathbf{S}_3\Big] \times \Big(\mathbf{e}_2 \star 1/2(\mathbf{e}_1+\mathbf{e}_3)\Big) \\ & \cup \Big[\Big((id \star \tau_3) \cup (\rho \star \tau_2) \cup (\rho^2 \star \tau_1)\Big) \setminus \mathbf{S}_3\Big] \times \Big(\mathbf{e}_3 \star 1/2(\mathbf{e}_1+\mathbf{e}_2)\Big) \\ & \cup \Big[\mathbf{B}_3 \setminus \Big(\{id, \rho, \rho^2\} \star \{\tau_1, \tau_2, \tau_3\}\Big)\Big] \times \{U\}, \end{aligned}$$

where on the last line here we use the fact that the join of those two finite point sets is the union of nine line segments that connect a point in the first set to a point in the second set.

Acknowledgement

We thank David DeGeorge for several enlightening comments.

References

[1] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 379–423, pp.623–656.
[2] C.E. Shannon, A mathematical theory of communication, in: IEEE Symposium on Information Theory at MIT, 50th anniversary ed., 1998.
[3] H. Minc, Nonnegative Matrices, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, New York, 1988..
[4] S. Eilenberg, N. Steenrod, Foundations of Algebraic Topology, Princeton University Press, Princeton, 1952.
[5] G.H. Hardy, J.E. Littlewood, G. Pólya, Inequalities, second ed., Cambridge University Press, Cambridge, 1952.
[6] C.P. Rourke, B.J. Sanderson, Introduction to Piecewise-Linear Topology, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 69, Springer-Verlag, New York, Berlin, Heidelberg, 1972.
[7] I. Bengtsson, E. Åsa, M. Kuś, W. Tadej, K. Życzkowski, Birkhoff's polytope and unistochastic matrices, N = 3 and N = 4, Commun. Math. Phys. 259 (2005) 307–324.